CHAPTER V

Experimental Results

This section illustrates the face database, the speed of the detector, and the experiments on real-world face images.

5.1 Face Database

The face and facial feature detection algorithms are applied to detect generic faces from several face image databases as follows:

1. ESSEX-face database [5] which contains head or head and shoulder pictures.
2. Purdue AR database [6] which contains face images with different facial expressions and occlusions under different light conditions.
3. MIT-CMU database [69] which contains gray-scale mugshot-style images.
4. Max Planck face database [7] which contains seven views of face images without hair.
5. UMIST face database [70] which consists of 564 images of 20 people, each covering a range of poses from profile to frontal views.
6. Private gallery which contains images from the World Wide Web, magazines, newspapers, and photographs.

(a) (b) (c)

Figure 5.1: Example of face images used for generating mean face template and training NVM detector. (a) Left-near-profile-view face images. (b) Frontal-view face images. (c) Right-near-profile-view face images.

A set of mean face templates for face detection algorithm is generated from gray images in ESSEX and Max Planck face databases, both of which contain face with single color background. Some face examples are shown in Figure 5.1. Since these face images have single color background, it is easy to extract face parameters from them. Hence, these images are also used in training the NVM. The test set for this study is obtained from the combination among those six databases. For the ESSEX face database, there are 20 faces varied in facial expression, scaling, and translation for each person, therefore only one image is selected at random for each person. Similar to the ESSEX database, 180 test images are randomly selected from the Purdue AR database under different light conditions and occlusions. For MIT-CMU database, the test set is originally separated. The Max Planck face database contains seven views of color face images. The gray version of color face images in this database is used for generating mean face template since it contains various views of black-background face images without hair. However, the actual run of face and facial feature detection algorithms also apply to all original color images in this database. For the UMIST face database, all images are tested for multi-view problem.

5.2 Speed of the Algorithms

The proposed method was carried out on Matlab V 6.0 with 1.4 Ghz Pentium 4 PC, generating a set of mean face templates from 200 faces of size 256×256 takes 61 seconds. The NVM consists of three neural visual networks for frontal, left-profile, and right-profile views. For each view, four facial features were located using four corresponding MLP networks. Training time for each MLP network was about 20 hours. The face detection algorithm can process an image of size 1280×1024 within 1 minute for each view. The facial feature detection algorithm takes about 15 seconds to locate all facial features in a face. This is 5 times slower than Rowley-Baluja-Kanade method [69] and 8 times faster than Schneiderman-Kanade method [71]. There are four main circumstances that affect the processing time as follows:

1. The multi-view face detection takes more time than single-view face detection.
2. It should be faster if the proposed method applied to integral image [55], a new image representation for very fast calculation, instead.
3. The parallel processing can be applied to improve the speed of face detection algorithm because each image is separated into subimages and they can be processed independently.
4. Matlab is somewhat slower than any other programming languages such as C++.

5.3 Combining Overlapping Face Regions

Overlapping detections usually occur around each face when length and width of face template are much larger than step size of scanning process, but the result should be one detection per face. To find the final result of face detection, postprocessing is applied

(a) (b)

Figure 5.2: Example of combining overlapping face regions. (a) Overlapping face regions. (b) Single detected face.

to combine overlapping detected faces into a single detected face. In this study, a conventional technique [55] is employed in postprocessing stage. The set of overlapping detected faces can be partitioned into disjoint subsets. Multiple detected faces are in the same subset if their boundary regions overlap. Finally, since there should be only one detected face for each partition, the four corners of the final boundary region are the average of the corners of all detected faces in the subset. In addition, reducing multiple detected faces into a single detected face decreases the number of false positives. Figure 5.2 shows the result of combining overlapping regions.

5.4 Experiments on Real-World Face Images

All color and gray images are converted to edge version. The normalized face size after face detection is 128 × 128 pixels. There are three sets of neural visual networks for multi-view faces. Each network consists of four MLP networks for detecting four essential features. There are seven nodes with respect to facial parameters in input layer and eighty nodes in hidden layer for each facial feature. The learning rate of the

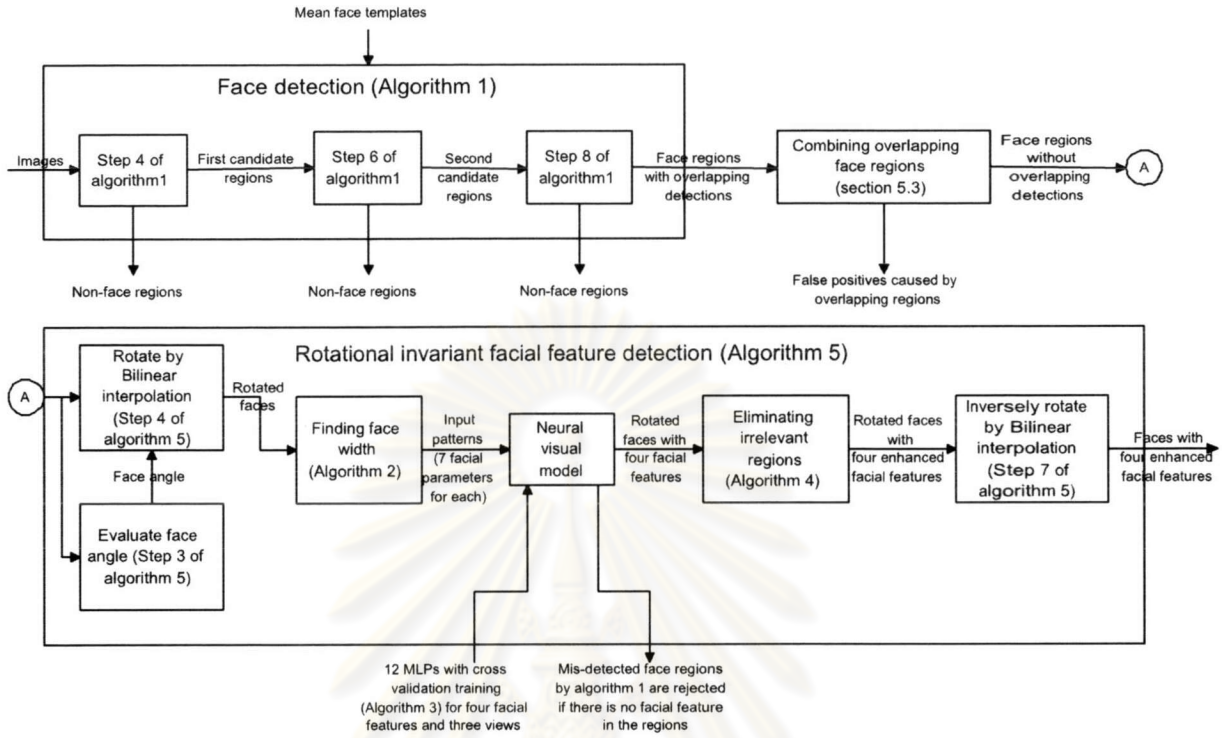
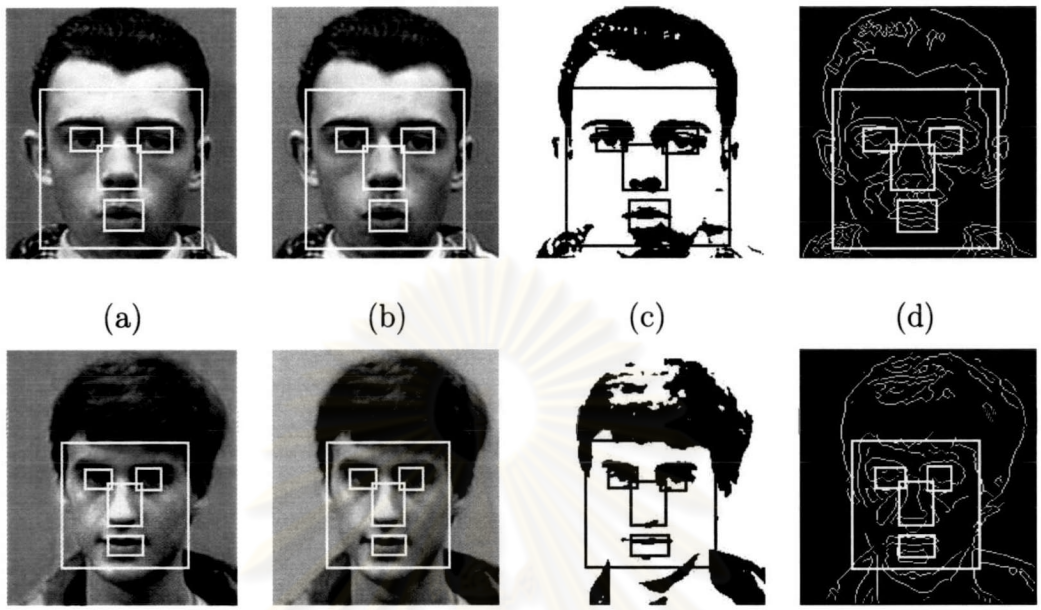


Figure 5.3: The overall process of our proposed face and facial feature detection.

BP learning algorithm is 0.05. The weights are updated using the delta rule without a momentum term in batch mode until the error is less than 0.001. The block diagram of the overall process is shown in Figure 5.3. The results can be categorized by the factors effective to the appearance of image as follows:

1. **Intensity.** The face and facial feature detection can be applied to images with any types of intensity. Most existing techniques are designed for only one type of image intensity. The techniques for color images [24,25] cannot be applied to gray, black-and-white, or edge images because the intensity information is not enough. Similar to the techniques for gray images [35,38], it is difficult to detect faces and facial features from black-and-white or edge images. Figure 5.4 shows the results.

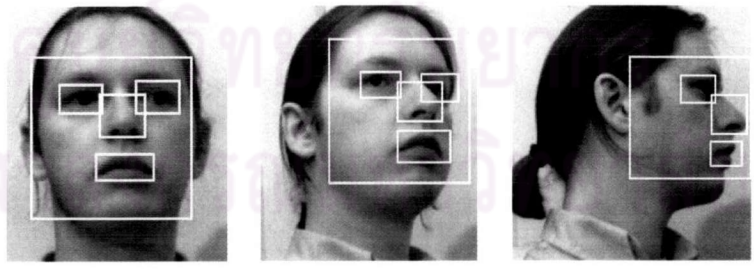
2. **Pose.** The multi-view face images are relative to camera-face pose, in which some facial features such as one eye may become partially (near-profile view) or



(a) (b) (c) (d)

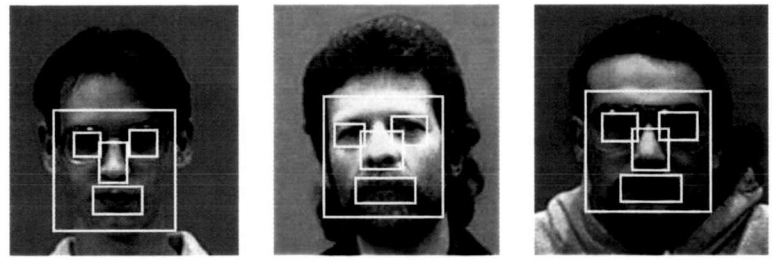
(e) (f) (g) (h)

Figure 5.4: Detection results of images with any kinds of intensity. (a) and (e) Color images. (b) and (f) Gray images. (c) and (g) Black-and-white images. (d) and (h) Edge images.



(a) (b) (c)

Figure 5.5: Multi-view face detection results. (a) Frontal view. (b) Right-near-profile view. (c) Right-profile view.



(a) (b) (c)

Figure 5.6: Detection results of face images with some structural components. (a) Face with glasses. (b) Face with beards and mustaches. (c) Face with glasses, beards, and mustaches.

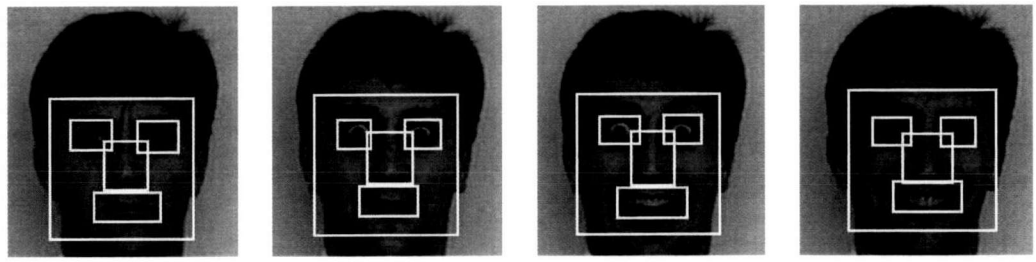
wholly occluded (profile view). Most face and facial feature detection methods are applicable only for frontal-view images. Figure 5.5 shows some detection results of multi-view face images.

3. **Structural components.** The face and facial features can be correctly detected from the face images with glasses, beards, and mustaches. Figure 5.6 shows the results.

4. **Facial expression.** This is an effective factor to the appearance of face. There are many types of facial expressions such as happy, sad, angry, etc. Figure 5.7 shows detection results of images with some facial expressions.

5. **Occlusion.** In this class, some facial features are occluded by an object such as sunglasses or scarf. The intensity information of the occluded features vanishes. Because the proposed algorithms use the position and face shape information, faces and facial features can be detected as shown in Figure 5.8.

6. **Image Rotation.** By rotational invariant facial feature detection algorithm, an angle of rotated face image is evaluated and features can be detected. Figure 5.9

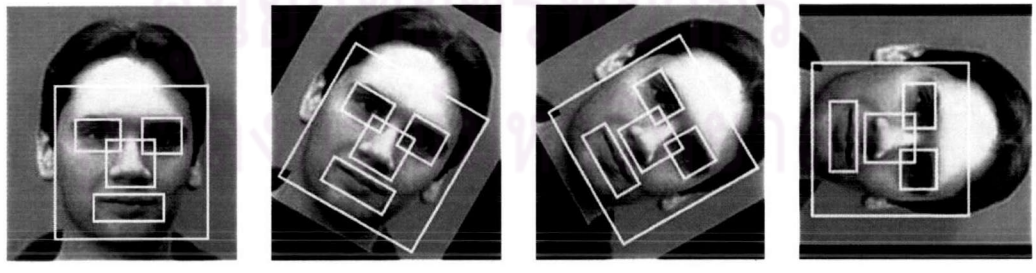


(a) (b) (c) (d)

Figure 5.7: Detection results of face images with facial expressions. (a) Angry face. (b) Excited face. (c) Very excited face. (d) Happy face.

(a) (b)

Figure 5.8: Detection results of face images with occlusions. (a) Face with sunglasses. (b) Face with scarf.



(a) (b) (c) (d)

Figure 5.9: Detection results of rotated face images. (a) 0° in clockwise direction. (b) 30° in clockwise direction. (c) 60° in clockwise direction. (d) 90° in clockwise direction.

(a) (b) (c)

Figure 5.10: Detection results of face images with lighting effect. (a) Left light source. (b) Front light source. (c) Right light source.

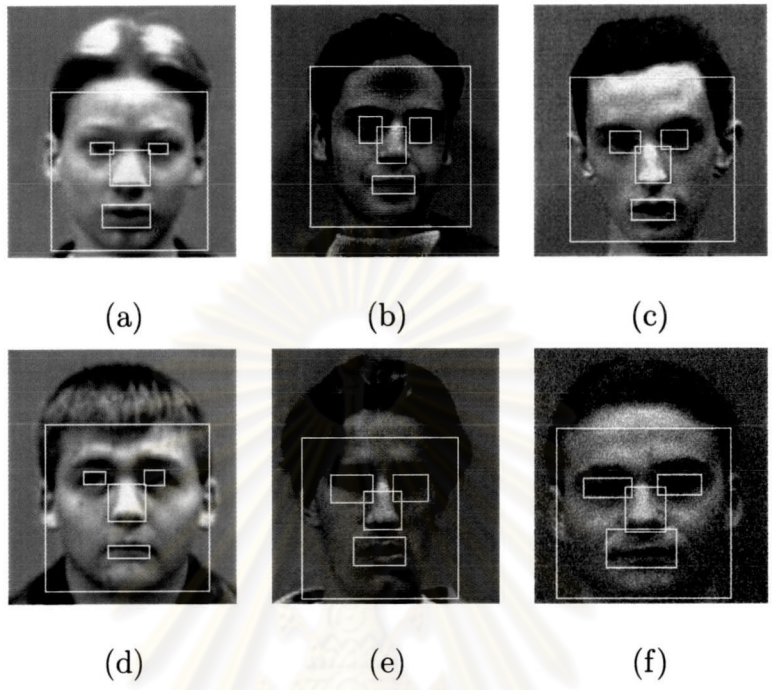
shows the detection results of rotated face images.

7. **Image Conditions.** Some conditions such as lighting effect and camera characteristics affect the appearance of a face when taking a photo. Figure 5.10 shows the lighting effects in three directions and the detection results.

8. **Poor quality.** In this class, faces and facial features in images with poor quality are detected. There are three subclasses in this category 1) blurry images obtained using blur filter, 2) distorted images by diffuse glow filter, and 3) images with 10% uniform or Gaussian noise. The results are shown in Figure 5.11.

9. **Unnatural intensity.** This class consists of cartoons, rendered images, and mosaic images. Their texture color is not natural as that in a human face image. It is difficult to detect faces and facial features by existing methods. Figure 5.12 shows some examples of unnatural-intensity images and the detection results.

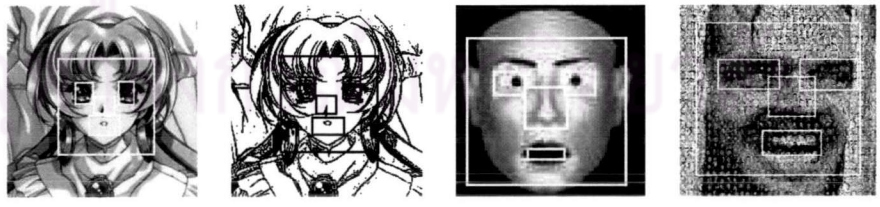
Generally, there are more than one factors simultaneously affecting the appearance of face image. Figures 5.13 and 5.14 show the detection results of such images. Some images contain many faces as in Figure 5.13(l) and Figures 5.14(d), 5.14(e), 5.14(g),



(a) (b) (c)

(d) (e) (f)

Figure 5.11: Poor-quality image detection results. (a) and (d) Blurry images. (b) and (e) Distorted images. (c) Image with 10% uniform noise. (e) Image with 10% Gaussian noise.



(a) (b)

Figure 5.12: Detection results of unnatural-intensity images. (a) Colored cartoon images. (b) Sketchy cartoon images. (c) Rendered images. (d) Mosaic images.

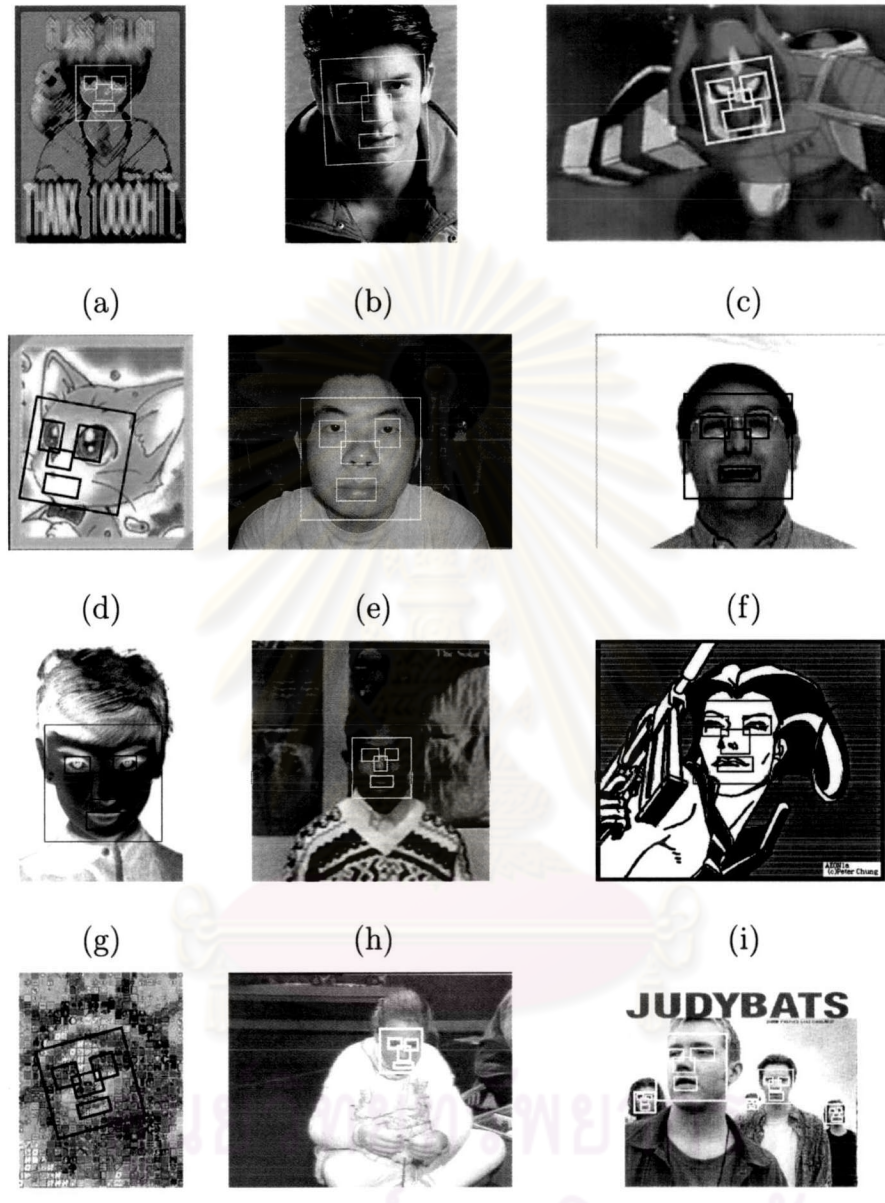


Figure 5.13: More results. (a) Gray-scale cartoon by colored pencil. (b) Rotated face with lighting effect. (c) Color cartoon with rotation. (d) Left-near-profile-view cartoon cat face. (e) Color face. (f) Gray face with glasses and facial expression. (g) Inverted black-and-white face like film negative. (h) Color dark-skin face. (i) Sketchy image. (j) Upper-left-near-frontal-view mosaic face. (k) Lower-frontal-view face. (l) Multi-view faces with size variation.

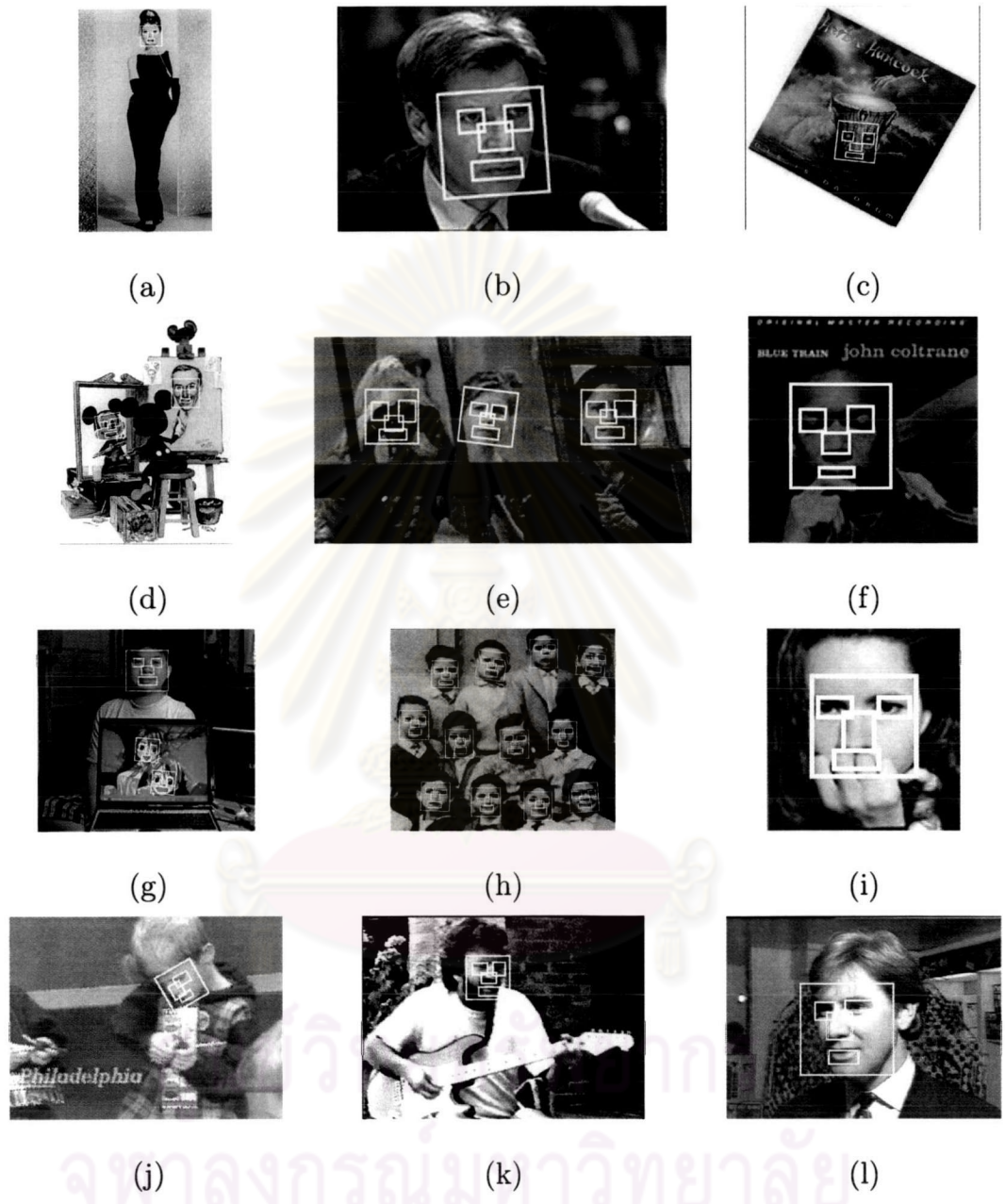


Figure 5.14: More results. (a) Gray image. (b) Right-near-frontal-view face with rotation. (c) Rotated unnatural face. (d) Cartoon and painted face with rotation and facial expression. (e) Multifaces with occlusion and rotation. (f) Gray dark-skin face with dark background. (g) Multifaces including cartoon faces. (h) Multifaces with rotation. (i) Occluded face. (j) Lower-left-profile-view face. (k) Right-near-frontal-view face with beards and poor quality. (l) Left-near-frontal-view face.

(a) (b)

Figure 5.15: The comparison of detection results. (a) Detection by our algorithms. The dog's face and facial features are also detected. (b) Detection by Bayesian discriminating feature method [38].

and 5.14(h). Some include rotated faces as in Figures 5.13(b)($-3°$), 5.13(c)($-10°$), 5.13(d)(10°), and 5.13(j)($-15°$) and Figures 5.14(b)($-4°$), 5.14(c)(6°), 5.14(d)($-10°$), 5.14(e)(7°), 5.14(g)(5°), 5.14(h)($-8°, -6°, 4°, 2°$), 5.14(j)($-30°$). Some contain non-frontal-view faces as in Figures 5.13(d), 5.13(i), 5.13(j), 5.13(k), and 5.13(l) and Figures 5.14(b), 5.14(j), 5.14(k), and 5.14(l). Some contain unnatural human faces as in Figures 5.13(a), 5.13(c), 5.13(d), 5.13(i), 5.13(j), Figures 5.14(c), 5.14(d), and 5.14(g). Some contain some occlusions as in Figures 5.14(e), 5.14(f), and 5.14(i), or structural components as in Figure 5.13(f)(glasses), and Figure 5.14(k)(mastaches and beards). Some contain lighting effect as in Figure 5.13(b). Some contain facial expressions as in Figure 5.13(f), Figures 5.14(c), 5.14(g), and 5.14(h). Some contain dark skins as in Figure 5.13(h) and Figure 5.14(f). Some appear as negative as in Figure 5.13(e). Moreover, dog's face can be also detected by the proposed algorithms as shown in Figure 5.15. Comparative of the face and facial feature detection results (including detection rates and number of false positives) are presented in Figure 5.16 and Table 5.1. To create

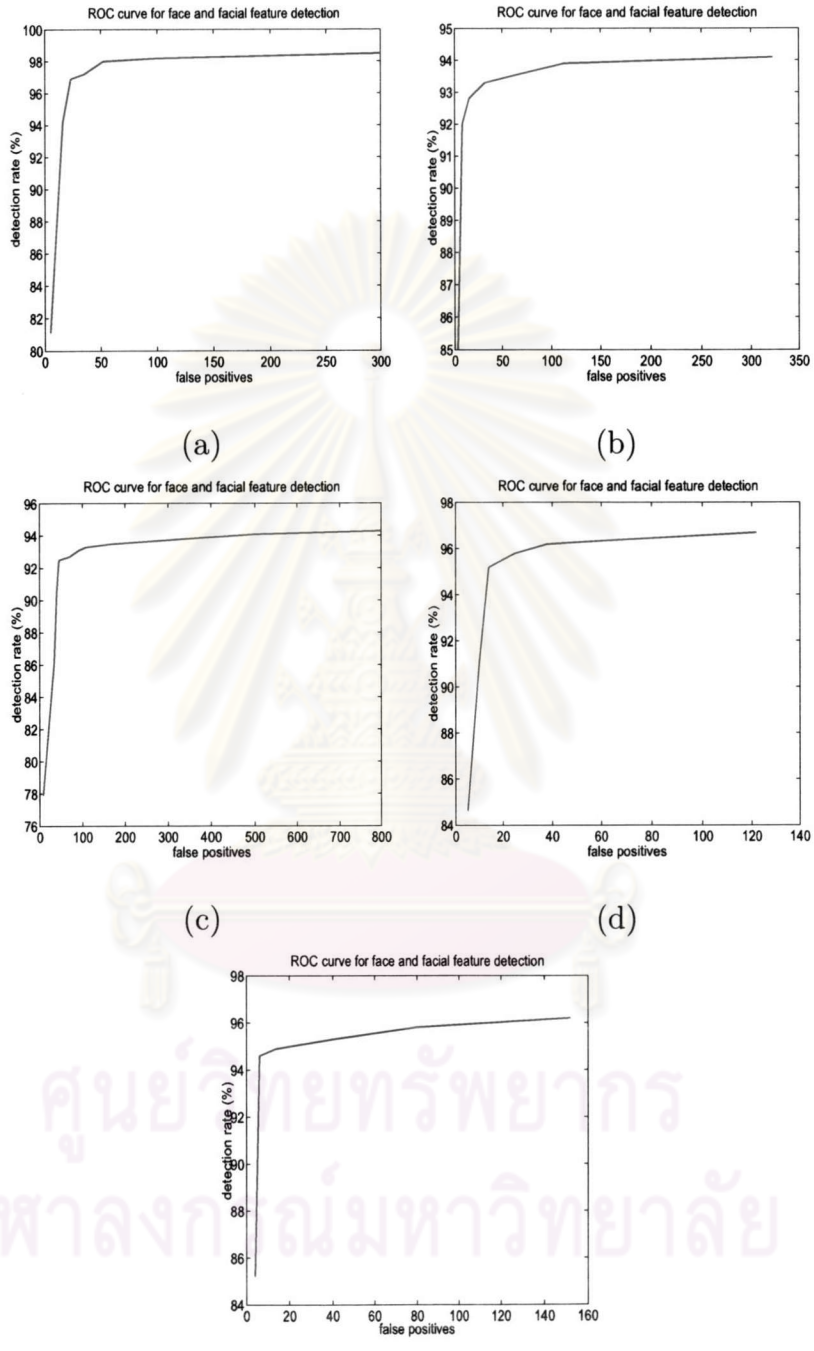


Figure 5.16: ROC curves for the proposed method on each database. (a) ESSEX-face database. (b) Purdue AR database. (c) MIT-CMU database. (d) Max Planck face database. (e) UMIST face database.

Table 5.1: Face and Facial Feature Detection Results of Each Database

Test set	ESSEX	AR	MIT-CMU	Max-Planck	UMIST
View Type	Frontal	Frontal	Frontal to near frontal	Frontal to profile	Frontal to profile
No. of images	397	180	130	1400	564
No. of faces	397	180	507	1400	564
No. of False positives	23	8	45	14	6
Detection rate (%)	96.88	92.31	92.48	95.23	94.61

the ROC curves, two thresholds in steps 6 and 8 in face detection algorithm (algorithm 1) and four thresholds of four MLP classifiers used for detecting the facial features are adjusted in the range between 0 and 1. Higher thresholds yield detector with fewer false positives and a lower detection rate. Lower thresholds yield detector with more false positives and a higher detection rate. Table 5.1 shows the best result for each database, i.e., single point on each ROC curve. A detected face is a correct detection if the detected locations of facial features and square block bounding a face are extracted with a small amount of tolerance, otherwise false positive is obtained. The detection rate is calculated from the ratio of the number of correct detections in a gallery to those of all human faces in the gallery. In the case of false positive, the error is caused by the edge regions that is similar to edge face in face detection stage, but it can be reduced in facial feature detection stage. For the AR [6], the Max Planck [7], and the UMIST [70] databases, the number of false positives is very low because each image has one face and single color background. Failure cases can be categorized into six major classes. The first failure class comes from the unusual locations of facial features or unnatural faces as depicted in Figure 5.17(a). The second class consists of the faces without face boundaries. Some

facial parameters such as face width are not able to be obtained by finding-face-width algorithm. An example of this class is shown in Figure 5.17(b) . The third class occurs when the the boundary regions of multiple faces are overlapped. A single detected face is computed by averaging the corners of all detected faces. This causes failure detection as shown in Figure 5.17(c). The fourth class comes from partial face, where face information is not enough for detection as the rightmost face in Figure 5.17(c). Subsequently, the fifth failure class is a set of very small images with low contrast such as some small faces in Figure 5.17(d). Finally, the last class consists of the face with incomplete thinned boundary. Since the face shape information is incorrectly obtained from the thinned boundary, the location of facial features may or may not be correctly detected as shown in Figure 5.17(e). Table 5.2 shows the effect from upper bound of number of white pixels in step 4 of face detection algorithm. The cofficient multiplied with γ, the number of white pixels in the edge face template, demonstrates the upper bound of number of white pixels. This bound affect the number of first candidate regions. Higher upper bound yields higher number of first candidate regions. In this study, the upper bound that give good implementation in the databases is 1.5γ. This value can handle the effect of thresholds from Canny edge detection. However, the upper bound is increased to 2γ for certainly covering other real-world face images excluded in the databases. Table 5.3 shows face detection rate and the number of false positives for proposed method as well as other existing methods. The Schneiderman-Kanade method [71] achieves 94.4% with 65 false positives. For the Rowley-Baluja-Kanade results [69], there are a number of different versions of detector tested yielding a number of different results, but their best detection rate is 90.1% with 167 false positives. For the Viola-Jones method [55], detection results are between 78.3% and 93.7% with a number of different values of false positives varying from 10 to 422. The detection rate with 65 false positives is selected. For the Xiao-Li-Zhang [34], the best detection rate is 92% with 135 false positives. For

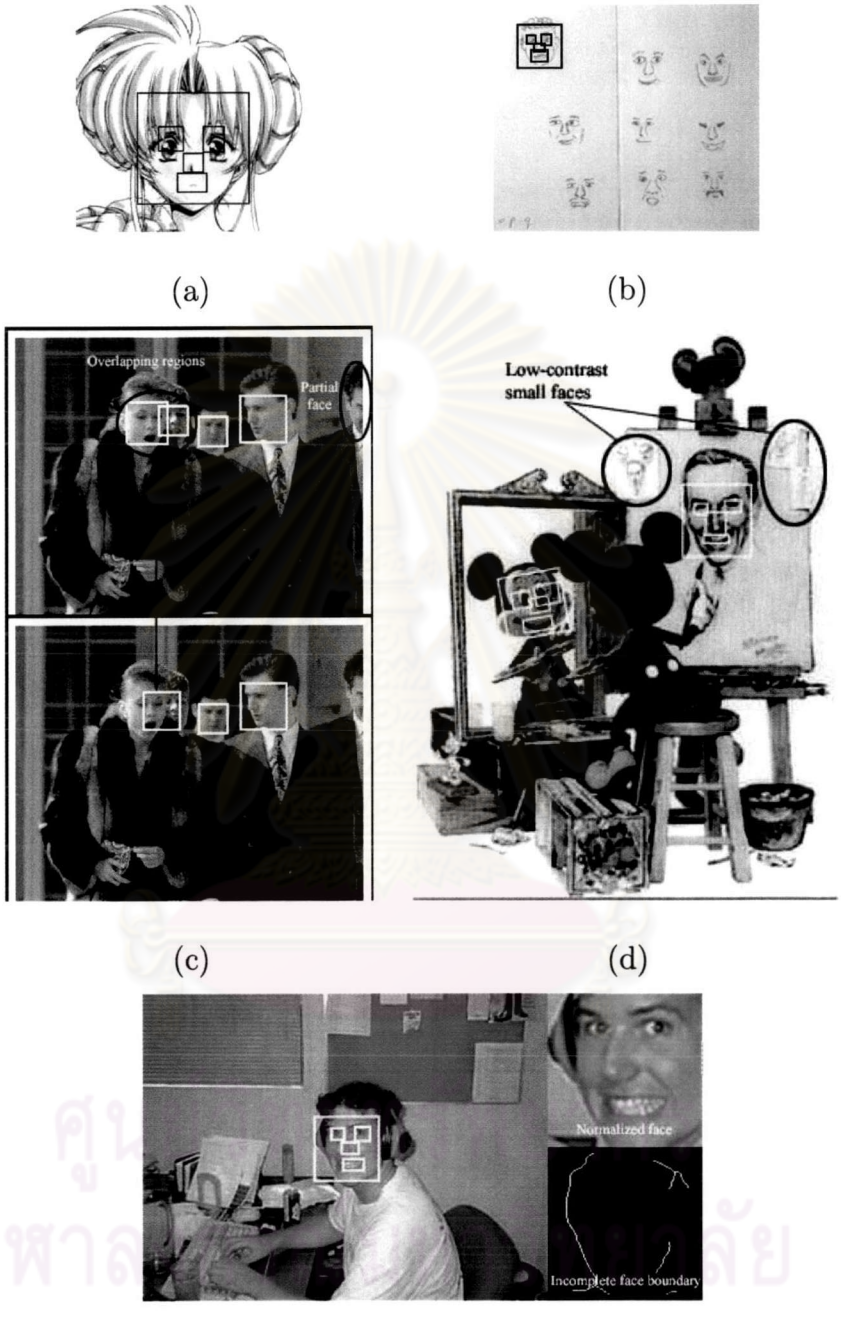
Table 5.2: Face Detection Rate Resulted from Varying Upper Bound of Number of White Pixels in Step 4 of Face Detection Algorithm on 20 Images Chosen at Random from the Databases

Upper Bound	Number of First Candidate Regions in 20 Images	Detection Rate(%)
γ	23920	75.6
1.25γ	65839	91.7
1.5γ	117512	96.2
2γ	165541	96.2

Table 5.3: Comparative Face Detection Performance of the Existing Methods and the Proposed Method on MIT-CMU Testset Containing 130 Images and 507 Faces

Method	Detection Rate(%)	False Detections
Proposed method	95.7	61
Schneiderman-Kanade [71]	94.4	65
Rowley-Baluja-Kanade [69]	90.1	167
Viola-Jones [55]	89.8	65
Xiao-Li-Zhang [34]	92.0	135
Huang-Shimizu-Hagihara-Kobatake [37]	86.0	53

Huang-Shimizu-Hagihara-Kobatake [37], the best classifier yields 86.0% detection rate with 53 false positives. The detection rate of proposed method is 95.7% with 61 false positives.



(a) (b) (c) (d) (e)

Figure 5.17: Failure cases. (a) Unusual eye locations. (b) Faces without face boundaries. (c) Overlapping boundary regions (upper image) causing a failure detected face (lower image) and partial face(rightmost face). (d) Very small faces with extremely low contrast. (e) Face with incomplete thinned boundary.